

Network Structure & Information Advantage: Structural Determinants of Access to Novel Information and Their Performance Implications

Sinan Aral
NYU Stern School & MIT Sloan School
Room: NE20-336, 3 Cambridge Center, Cambridge, MA 02142
sinana@mit.edu

(Research in collaboration with: Erik Brynjolfsson & Marshall Van Alstyne)

We explore new methods of mining and analyzing organizational information flows and individual information worker productivity. We extract and link corporate email data to accounting data on employee performance to test relationships between social network structures, information structures, and the performance and productivity of individual employees. Specifically, we investigate which network structures influence access to diverse and novel information, and whether these relationships explain performance in information intensive work.

Keywords: Social Networks, Information Content, Information Diversity, Productivity, Econometrics.

History: Presented at the Workshop on Information Systems Economics 2006, Chicago IL.

Status: A full length draft is complete and can be distributed at the conference.

Extended Abstract Submitted to:

*The 2007 Winter Conference on Business Intelligence,
University of Utah*

Introduction

A growing body of evidence links the structural properties of individuals' and groups' networked relationships to various dimensions of economic performance. However, the mechanisms driving this linkage, thought to be related to the value of the information flowing between connected actors, are typically inferred, and rarely empirically demonstrated. Comprehensive theories of the structure-performance relationship require a more thorough examination of the intermediate mechanisms through which social structure affects economic advantage. The strategy of this paper is to narrowly examine one of these mechanisms – the relationship between *network structures* and *information structures* – in some detail.

We build hypotheses linking two key aspects of network structure – network size and network diversity – to the distribution of novel information among actors. Novel information is thought to be valuable due to its local scarcity. Actors with scarce, novel information in a given network neighborhood are better positioned to broker opportunities, use information as a commodity, or apply information to problems that are intractable given local knowledge. We test the implications of our theory using empirical evidence from a ten month panel of email communication patterns, message content, and performance among information workers in a medium sized executive recruiting firm.

Measuring Social Networks in Email Data

We measure social networks by analyzing email data. Network size is simply the number of contacts in an individual's email network. Network diversity describes the degree to which contacts are structurally 'non-redundant' in both first order and second order contacts as shown in Figure 1.

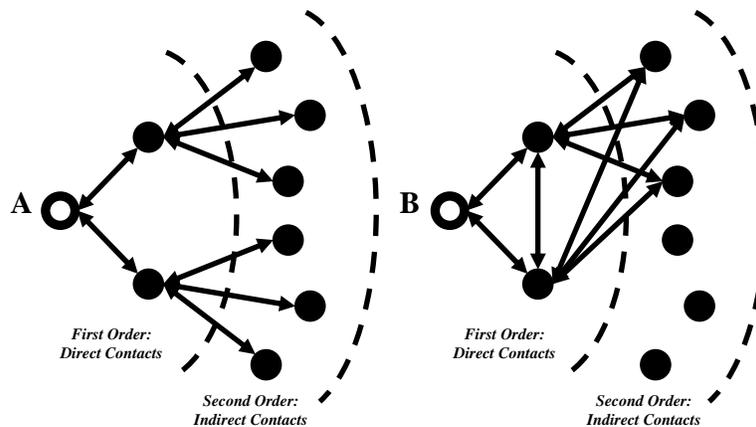


Figure 1. Structurally diverse networks are low in a) cohesion and b) structural equivalence. Cohesion measures the extent to which an actor's direct contacts are connected to each other. Structural equivalence measures the degree to which an actor's contacts are themselves connected to the same people. A has two unconnected contacts with no structural equivalence, while B has two redundant contacts that are connected and maximally structurally equivalent.

Individuals who are in contact, or whose contacts share contacts, are likely to share information and be aware of the same opportunities, ideas and expertise, connecting the focal actor directly and indirectly to redundant sources of information. We hypothesize the network diversity should impact productivity and performance in information intensive work by providing access to novel and non-redundant information. We test this hypothesis using data on the revenues generated, projects completed and average project duration of executive recruiters, controlling for

demography, human capital, total communication volume, unobserved characteristics of recruiters and temporal variation.

Modeling & Measuring Topics in Email: A Vector Space Model of Communication Content

We model information diversity and novel information using a Vector Space Model of email content. In our model, each email is represented as a multidimensional ‘topic vector’ whose elements are the frequencies of keywords in the email. For example, an email about pets might include two mentions of the word “dog,” two of “cat,” and three of “veterinarian;” while an email about econometrics might include three uses of the word “variance,” two of “specification,” and three of “heteroskedasticity.” The relative topic similarity of two emails are then assessed by the degree to which their topic vectors converge (point toward the same topics in vector space) or diverge (point in orthogonal directions in vector space) using several different measures of vector distance, such as the cosine similarity of topic vectors.

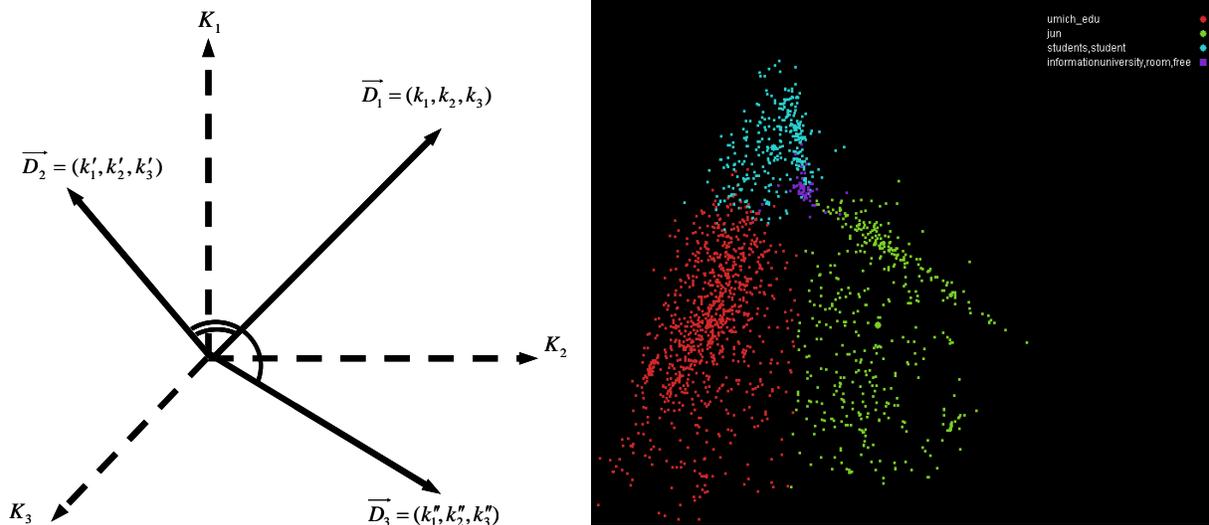


Figure 3. A three dimensional Vector Space Model of three documents is shown on the left. A Vector Space Model with emails clustered along three dimensions is shown on the right.¹

Findings & Conclusion

Our findings indicate that: (1) the total amount of novel information and the diversity of information flowing to actors are increasing in actors’ network size and network diversity, while (2) the marginal increase in information diversity is decreasing in network size. We also find evidence of a fundamental tradeoff between network size and network diversity. Part of the explanation for the decreasing marginal contribution of network size to information diversity is that (3) network diversity is increasing in network size, but with diminishing marginal returns. As actors establish relationships with a finite set of possible contacts in an organization, the probability that a marginal relationship will be non-redundant, and provide access to novel information, decreases as possible alters in the network are exhausted. We also find that (4) network diversity contributes to performance even when controlling for its positive performance effects through access to novel information, suggesting additional benefits to network diversity beyond those conferred through information advantage. Surprisingly, (5) traditional demographic and human capital variables (e.g. age, gender, industry experience, education) have little effect on

¹ The image on the right was created using IBM’s eClassifier tool.

access to diverse information, highlighting the importance of network structure for information advantage.

Our results represent some of the first empirical evidence on the relationship between network structure and information content, and our methods for analyzing network structure and information content in email data can be replicated in other settings, opening a new line of inquiry into the relationship between network structure and information advantage.

References

(The following paper was presented at last year's IS Winter Conference).

Aral, S., Brynjolfsson, E., & Van Alstyne, M. (2006). "Information, Technology and Information Worker Productivity: Task Level Evidence." Proceedings of the 27th Annual International Conference on Information Systems, Milwaukee, WI.